

**CovidData-19.** Una experiencia de recogida y utilización de datos durante durante la crisis de la pandemia.

# GRUPO DE INVESTIGACION

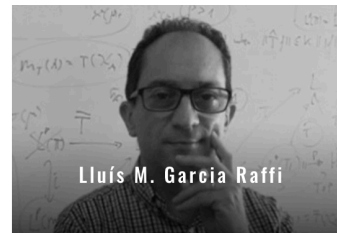
**MAD $\Psi$**  | MATHEMATICS  
& APPLICATIONS  
TO DATA & PHYSICS



José Manuel Calabuig



Enrique A. Sánchez Pérez



Lluís M. García Raffi



Antonia Ferrer Sapena



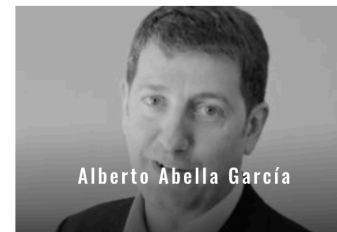
Fernanda Peset Mancebo



Rafael Aleixandre Benavent



Hervé Falciani



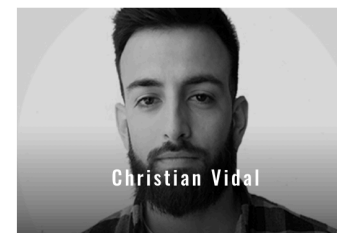
Alberto Abella García



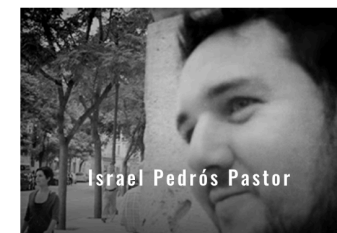
Juan Sebastian Bejarano  
Ballen



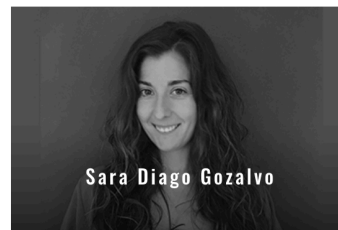
M. Isabel Sánchez Del Toro



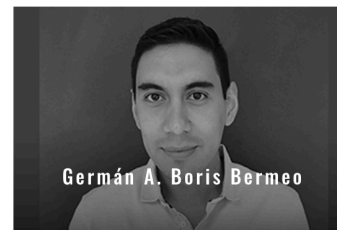
Christian Vidal



Israel Pedrós Pastor



Sara Diago Gozalvo



Germán A. Boris Bermeo



Ezgi Erdogan



Eduardo Jiménez Fernández



- Debido a la tensión en los sistemas sanitarios y a las medidas de emergencia del confinamiento e hibernación de la economía.

***DATAFILIA***



Photo by Luis Villasmil on Unsplash

# En la literatura científica

- Publicaciones relacionadas con encontrar una solución a la Covid en el área sanitaria y de la aplicación de inteligencia artificial a los datos.
- Aspectos más sociales, centrados en la disponibilidad de datos fiables y abiertos, no están siendo suficientemente atendidos.

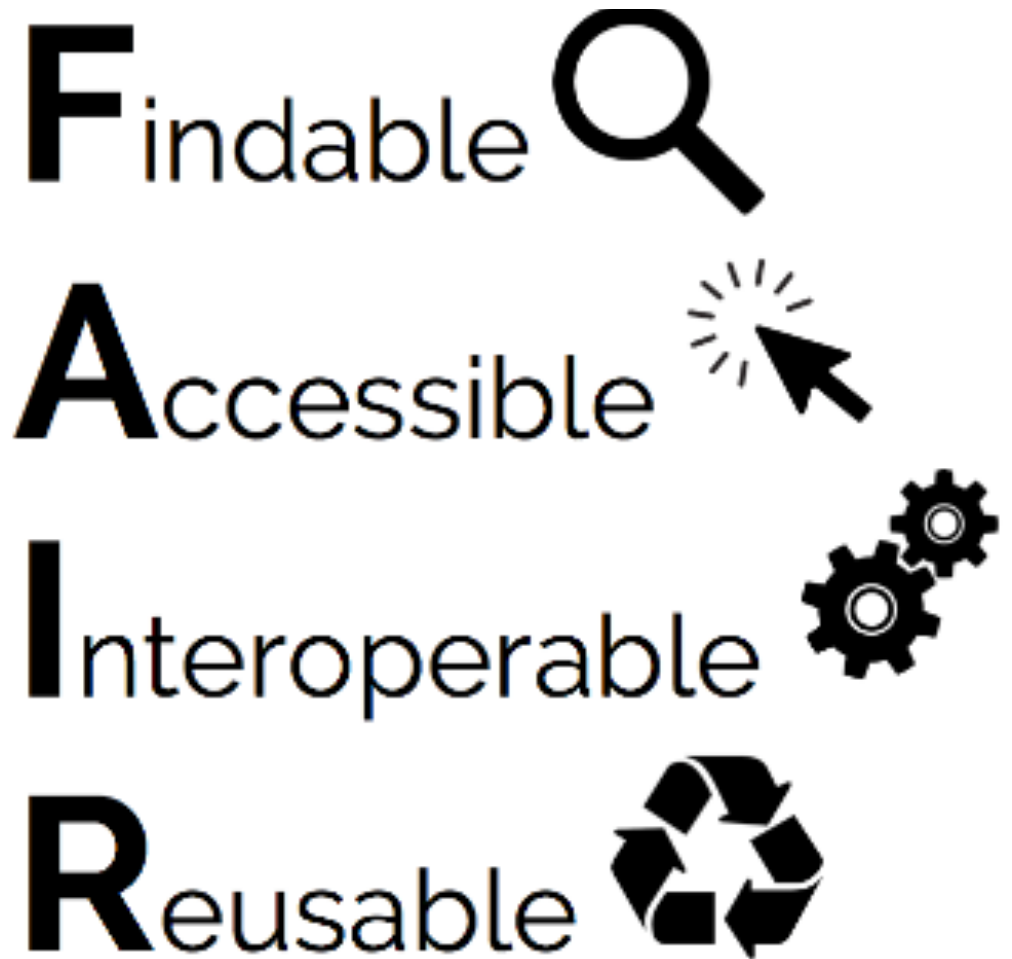


Photo by Trust "Tru" Katsande on Unsplash



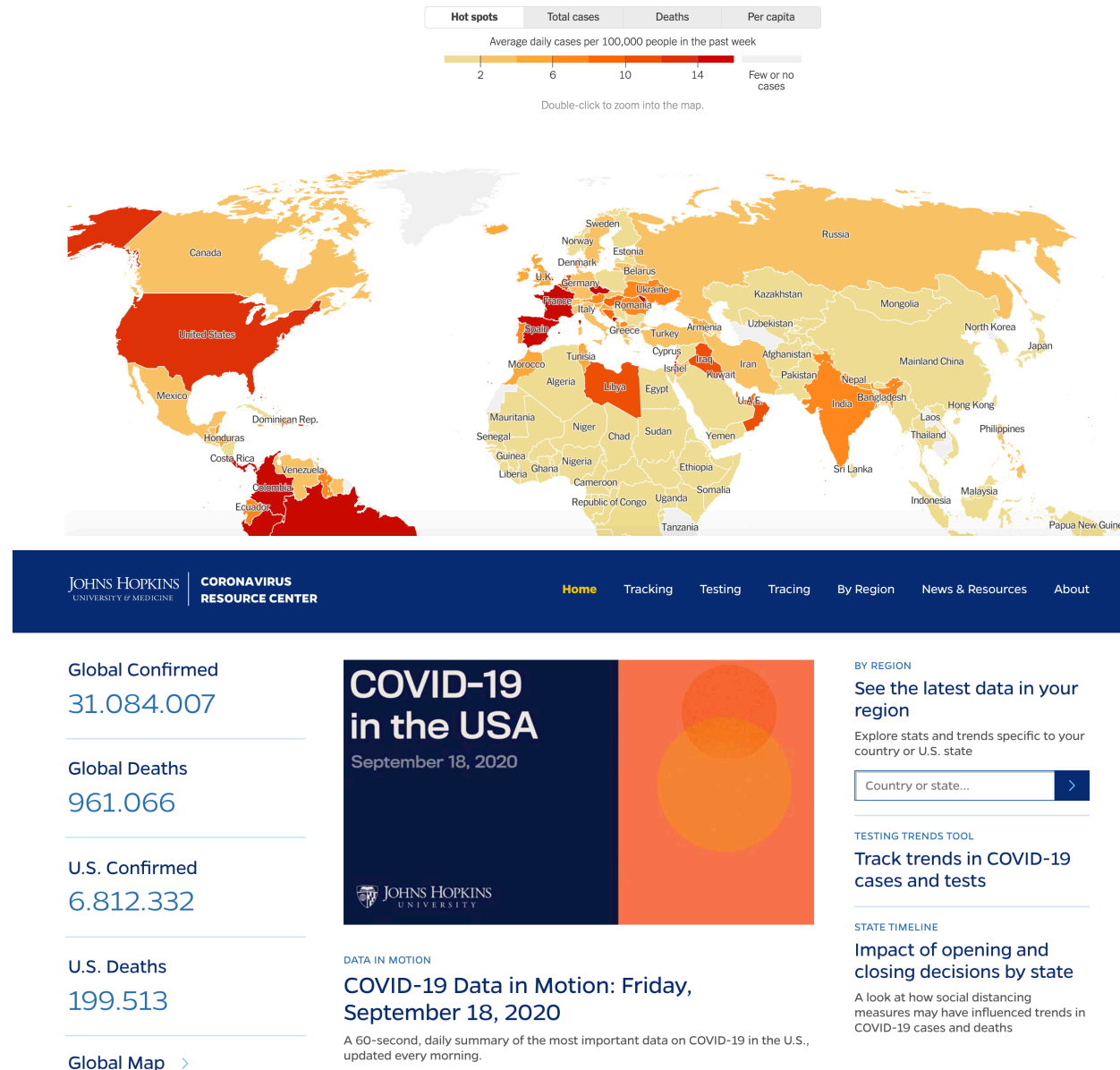
# Principios FAIR

- No se encuentran en los datos de infectados, fallecidos, ingresos UCI y recuperados, complementados con datos demográficos y topográficos o número de camas, de mascarillas...
- Los datos que se proporcionan son: dispersos, poco homogéneos en su forma de acceso, y con un nivel de interoperabilidad y capacidad de ser reutilizados escasa.



# Portales web

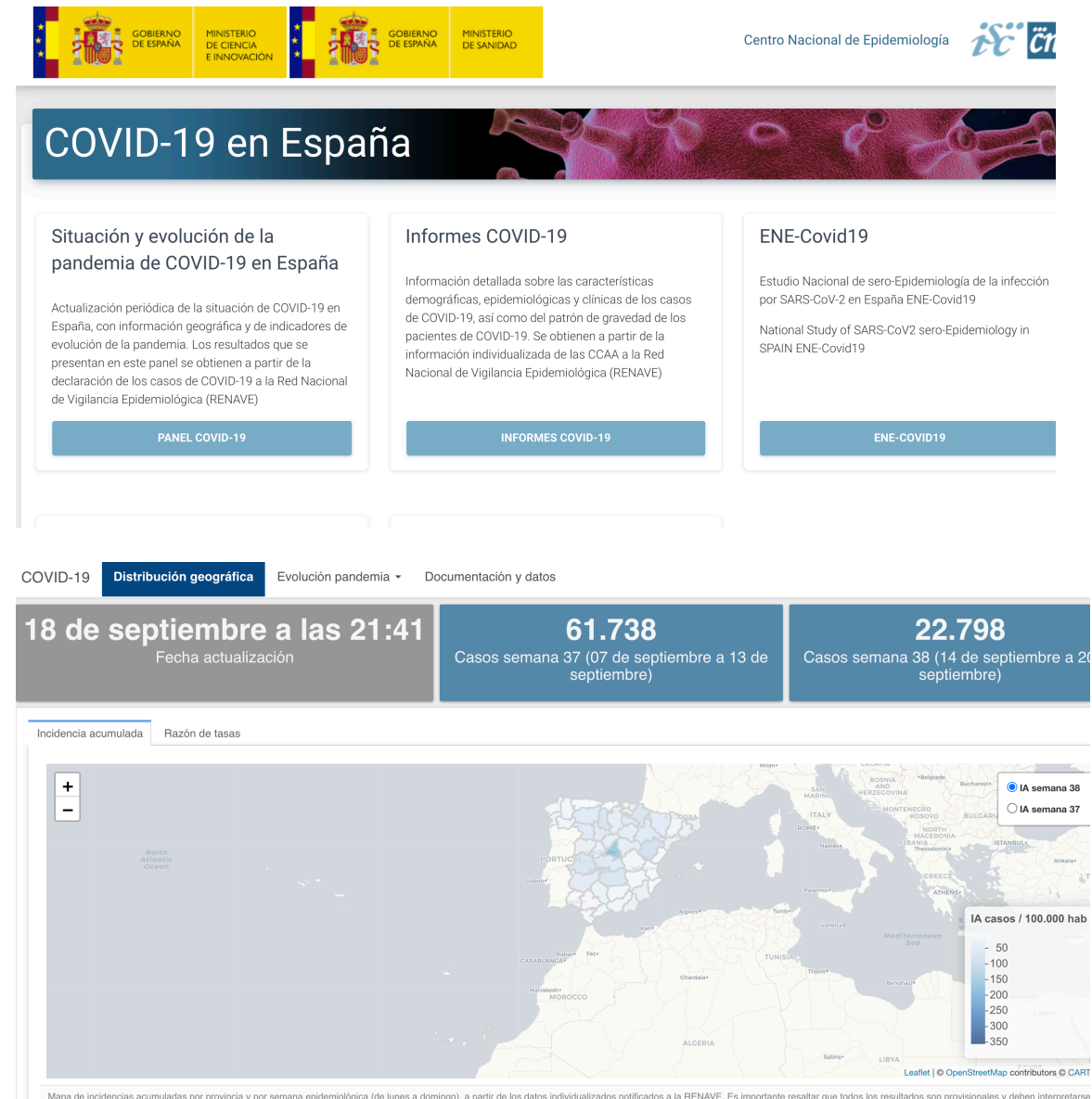
- Prácticamente todos incluyen: infectados, fallecidos y curados por distribución geográfica o de género.
- Las visualizaciones pueden incluir mapas, curvas de contagiados, o la proporción de éstos y fallecidos por población.





# Los datos en España

- Declara el estado de alarma el **14 de marzo de 2020** el Ministerio asumió la coordinación. Las competencias en sanidad están descentralizadas en 17 comunidades y 2 ciudades autónomas.
- Las CCAA deben reportar a la Red Nacional de Vigilancia Epidemiológica (Renave) los datos de los centros sanitarios sobre:
  - casos confirmados (acumulado)
  - casos de personal sanitario
  - hospitalizaciones
  - ingresos en UCI
  - recuperados
  - fallecidos.



# Los datos en España

## Problema

El problema más relevante es ***no haber determinado desde el inicio el método de recogida de información***: qué es lo que abarca cada uno de los indicadores de la muestra, y cómo y cuándo debe ser recogido.

**Datadista** recoge en su repositorio de GitHub cada una de las notificaciones asociadas al Ministerio de Sanidad que señalan los aspectos que a lo largo de las semanas han podido producir incidencias en la recogida de datos o notas aclaratorias sobre los mismos

### DATADISTA

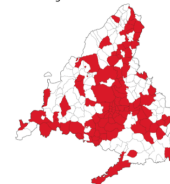
#### CRISIS DEL CORONAVIRUS

ZBS con confinamiento parcial acordado por Madrid



ZBS con una incidencia acumulada los últimos 14 días entre 810 y 1.883

Municipios que deberían ser confinados según el Ministerio de Sanidad



Madrid capital y los municipios con una incidencia acumulada los últimos 14 días > 500

Unidad de convivencia	Escala de incrementos	Renta anual	Rental mensual (12 pagas)
Un adulto solo	5,538	5.538	461,5
Un adulto y un menor	1,52	6.417,76	701,48
Un adulto y dos menores	1,82	10.079,16	839,93
Un adulto y tres o más menores	2,12	11.740,56	978,38
Dos adultos	1,3	7.199,4	599,95
Dos adultos y un menor	1,6	8.860,8	738,4
Dos adultos y dos menores	1,9	10.522,2	876,85
Dos adultos y tres o más menores	2,2	12.183,6	1.015,3
Tres adultos	1,6	8.860,8	738,4
Tres adultos y un menor	1,9	10.522,2	876,85
Tres adultos y dos o más niños	2,2	12.183,6	1.015,3
Cuatro adultos	1,9	10.522,2	876,85
Cuatro adultos y un niño	2,2	12.183,6	1.015,3
Otros	2,2	12.183,6	1.015,3

Madrid desoye a Sanidad y sumará al

ACTUALIZADO | Plazos, requisitos y cuantías



# Los datos en España

## Ministerio de Sanidad

- Información casi diaria pormenorizada con anotaciones correctoras.
- Primera nota 26 marzo.



*“Los casos confirmados no provienen de la suma de pacientes hospitalizados, curados y fallecidos, ya que no son excluyentes. Pacientes fallecidos y curados pueden haber precisado hospitalización y por tanto computar en ambos grupos. Los pacientes que han precisado UCI también computan en los pacientes que han requerido hospitalización”.*

# Los datos en España

## Ministerio de Sanidad

- Las notas aclaratorias del Ministerio de Sanidad incluyen un criterio diferenciador que algunas CCAA han incluido en su recogida de datos y que ha cambiado el modelo de entrega.
- Es el caso de **la prevalencia en casos de hospitalizados y en UCI** en el período del 2 al 7 de abril. La prevalencia **no son los casos acumulados a día de notificación** sino sólo los producidos en el mismo día.





# Los datos en España

## Problema

Las primeras CCAA en presentar estos datos de prevalencia en hospitalizados y en UCI fueron **Madrid, Castilla La Mancha y Comunitat Valenciana**.

El 3 de abril se añadieron **Castilla y León** a los datos de prevalencia de hospitalizados y **Galicia** a los de UCI.

El día 7 de abril **Castilla y León** sólo proporcionó los casos de prevalencia de UCI y el Ministerio de Sanidad siguió publicando los datos acumulados de hospitalizados y de UCI.



Comunidad de Madrid



Castilla\_La Mancha



Comunidad Valenciana



Castilla-León



Comunidad Gallega

# Los datos en España

## Problema

Consecuencia de estas incidencias concluyó el **8 de abril**: el Ministerio de Sanidad anunció que dejaba de publicar el sumatorio de todas las personas que han requerido hospitalización o ingreso en UCI porque los datos no reflejaban el total de personas que habían sido hospitalizadas o ingresadas en UCI a lo largo del período de notificación

**Los datos** que presentaron las CCAA mencionadas solo eran de **prevalencia**.



Photo by Pawe Czerwiski on Unsplash



# Los datos en España

## LÍNEA TEMPORAL DE INCIDENCIAS EN LA RECOGIDA DE DATOS DEL MINISTERIO DE SANIDAD

1

**26 MARZO**

Confirmados no son la suma de hospitalizados, fallecidos y altas

2

**2 ABRIL**

Hospitalizados en Castilla La Mancha, Comunitat Valenciana y Madrid son de **prevalencia**. En Madrid también los de UCI

3

**3 ABRIL**

Se añade Castilla y León a los datos de **prevalencia** en hospitalizados y Comunitat Valenciana, Galicia y Castilla La Mancha a los de UCI

4

**7 ABRIL**

Castilla y León da los datos de **prevalencia** sólo de los de UCI. Se sigue publicando el total de acumulados de hospitalizados y de UCI

5

**8 ABRIL**

Se deja de publicar el total de acumulados de hospitalizados y UCI

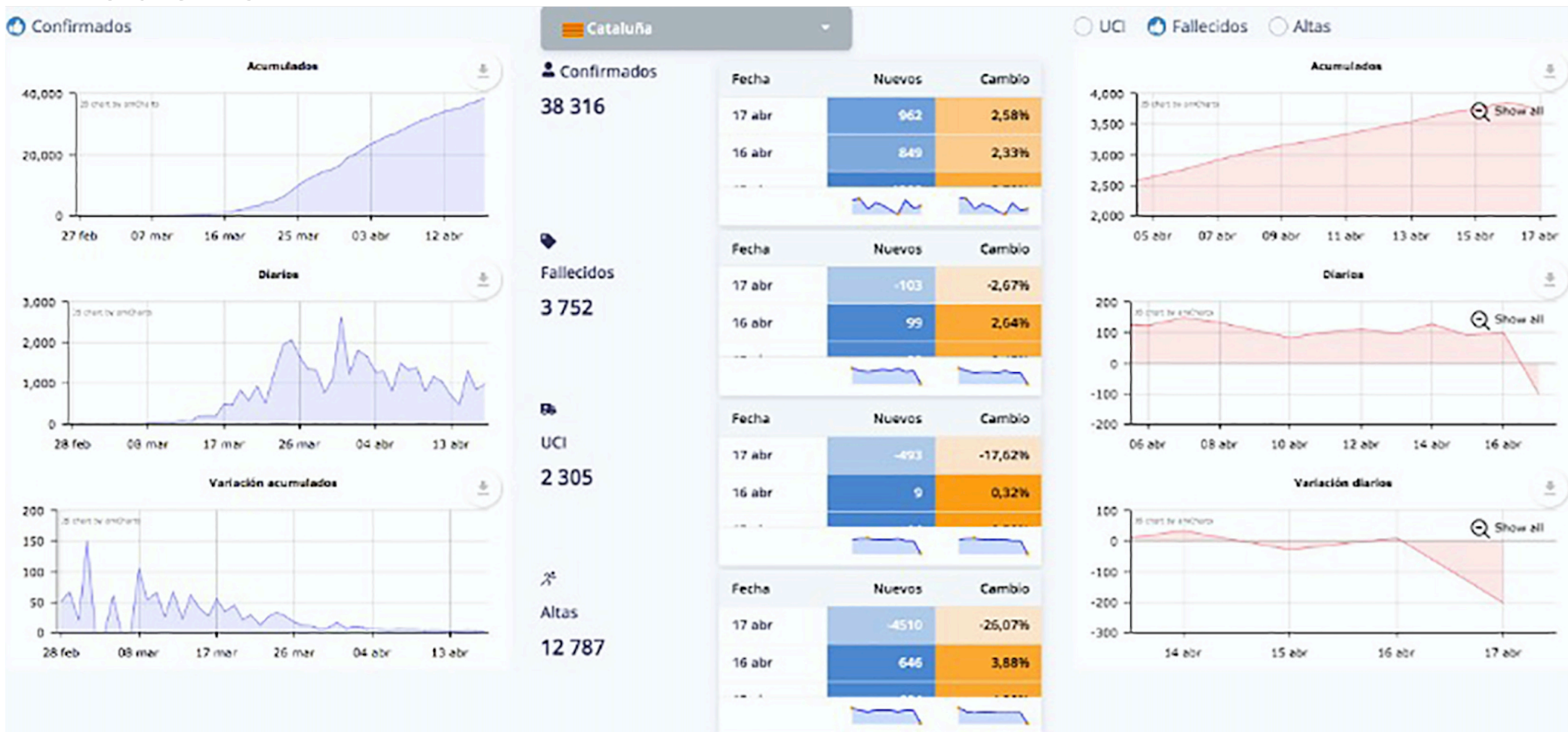
6

**15 ABRIL**

Andalucía incluye asintomáticos entre los confirmados

# Los datos en España

## Problema



# CovidData-19

## Recogida de datos

1. Las CCAA completan una encuesta individualizada para cada uno los casos según accedían a la información.
  2. La encuesta que incluía información clínico-epidemiológica consensuada y que había sido aprobada por la Ponencia de Alertas y Planes de Preparación y Respuesta y la Renave la notificaban mediante la plataforma informática SiViES, que gestiona el Centro Nacional de Epidemiología-Instituto de Salud Carlos III.
- Muchas CCAA crean sus propios portales:
    - En unos casos en los portales de datos abiertos
    - En otros casos portales propios del Covid, actualmente no funcionan la mayoría.

# **CovidData-19**

## **Obstáculos de los datos**

1. Los datos del Ministerio están en PDF, hay que convertirlos en reutilizables.
2. Diferencias horarias en la recogida, culmen 17 de abril en las críticas en el recuento de fallecidos.
3. Desde el inicio algunas comunidades no presentaban la misma información, ni los datos seguían la misma estructura. Por ello, no describen la misma realidad y la evolución, ni son fácilmente comparables; la cantidad de anotaciones mencionadas obligan a revisar y recalcular datos.

## **Providencialdata19**



# CovidData-19

## Providencialdata19. Registro de incoherencias

- **Comunidad Canaria** el 06/04/2020 los datos consolidados del día 05/04/2020 a las 20:00h son casos registrados en función del **lugar de declaración y no de residencia.**
  - Con este detalle el número de casos nuevos llega a ser negativo en algunas islas.



**Comunidad  
Autónoma  
de Canarias**

# CovidData-19

## Providencialdata19. Registro de incoherencias

Figuran los datos de las personas a las que se les ha practicado PCR y han dado positivo.



[https://es.wikipedia.org/wiki/Junta\\_de\\_Castilla\\_y\\_Le%C3%B3n#/media/](https://es.wikipedia.org/wiki/Junta_de_Castilla_y_Le%C3%B3n#/media/)

Archivo:Logotipo\_de\_la\_Junta\_de\_Castilla\_y\_Le%C3%B3n.svg

# CovidData-19

## Providencialdata19. Registro de incoherencias

Ofrece los datos totales diarios, pero no los acumulados. En las defunciones no se indica el municipio y tampoco se indica si son sólo los hospitalarios o si tienen también en cuenta los fallecidos en residencias, en casa...



**Generalitat  
de Catalunya**

# CovidData-19

## Providencialdata19. Registro de incoherencias:

No queda claro si el número que aparece es de activos o acumulados.

Las cifras de ingresos hospitalarios, fallecimientos, altas, UCIs y casos domiciliarios son globales para la Comunidad Autónoma y no se desglosan ni por provincias ni por distrito sanitario.

Hay discrepancias entre lo publicado por el Ministerio y la publicado por la Xunta.



**XUNTA  
DE GALICIA**



# CovidData-19

## Providencialdata19. Registro de incoherencias:

En Euskadi tampoco figuran los datos de fallecidos por provincia. Existen también discrepancias entre lo que algunos gobiernos autonómicos publican en Twitter y lo que figura en sus informes diarios.

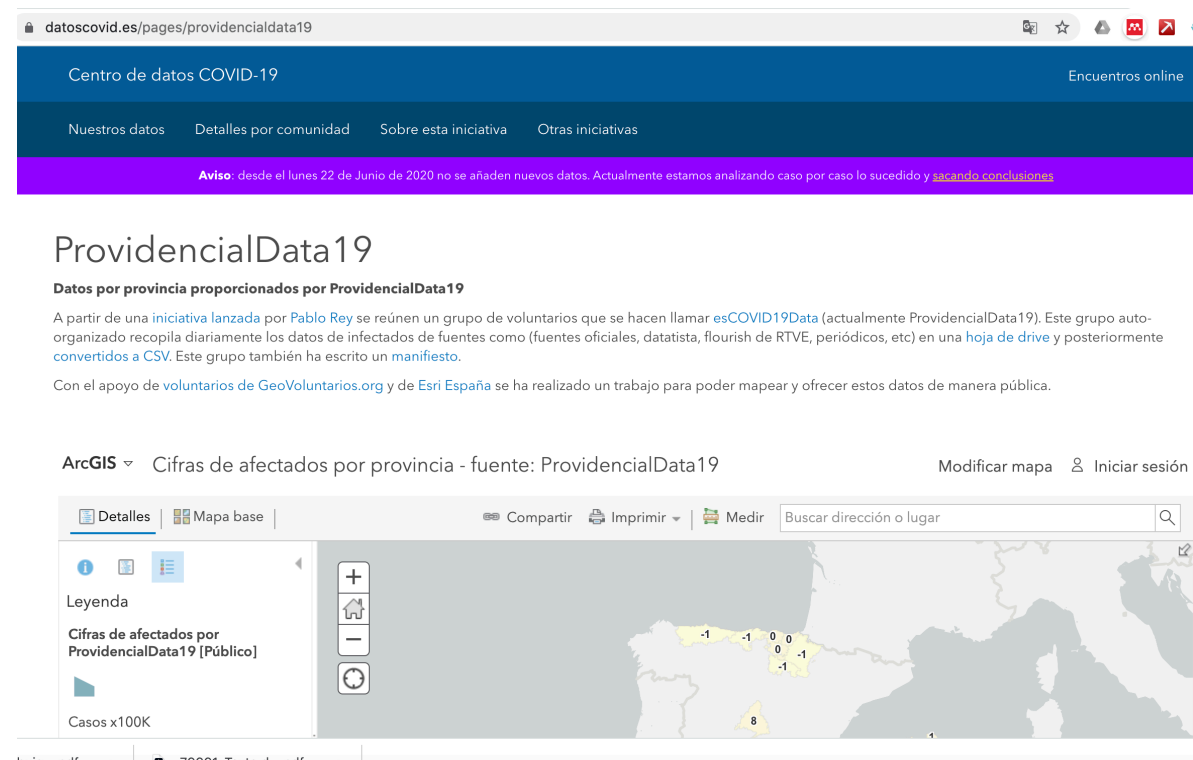


# CovidData-19

## Providencialdata19

Existe fragilidad de la confianza en los datos en los momentos de crisis.

Estamos asistiendo a una evolución desde la **datafilia** detectada en un principio a una **datafobia** en la que se descalifica de forma absoluta la validez de los datos de forma absoluta la validez de los datos de forma absoluta la validez de los datos



<https://www.datoscovid.es/pages/providencialdata19>

# CovidData-19

## Providencialdata19. Conclusiones

### Recomendaciones sobre normas protocolos y buenas prácticas:

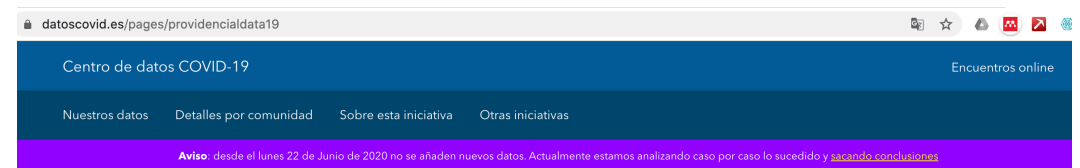
Los datos siempre tengan que ser facilitados directamente por fuentes oficiales

Todos los territorios deberían publicar los mismos datos, durante el mismo periodo, la misma frecuencia, siguiendo los mismo criterios y al máximo nivel de desagregación territorial posible (garantizando la anonimidad de los datos)

Se debe facilitar un fichero de metadatos

Actualizar los datos históricos cuando exista una agregación de datos y se cambie el criterio de contabilización....

<https://www.datoscovid.es/pages/sobre-la-iniciativa>



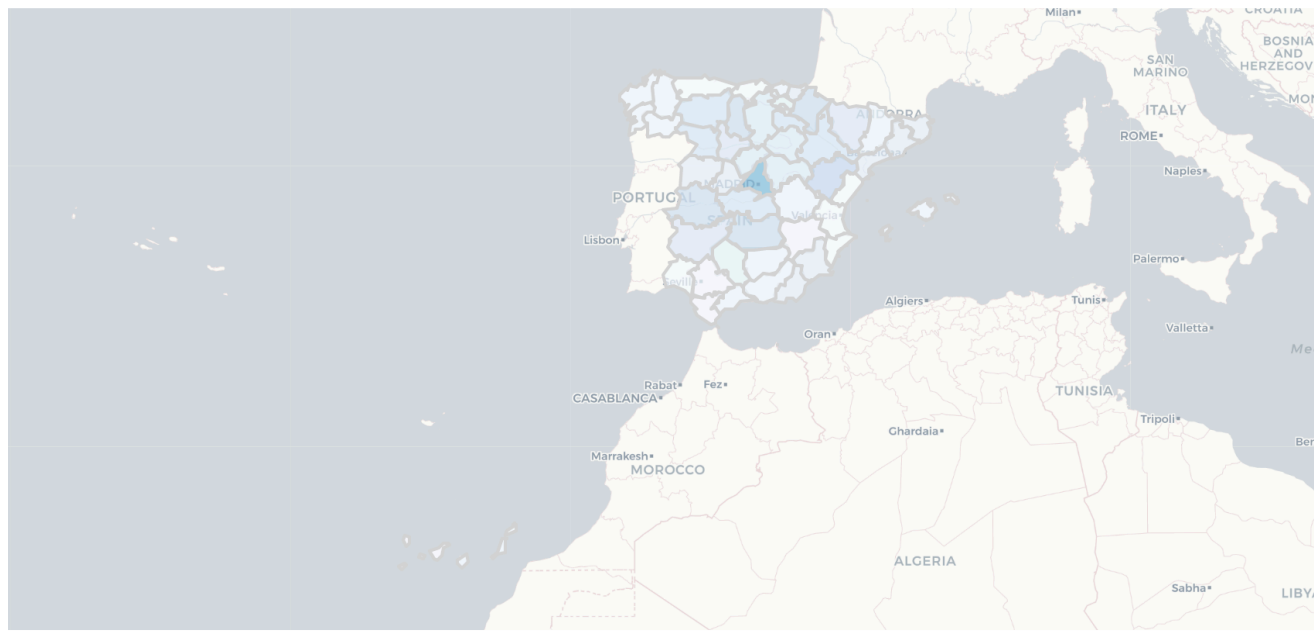
### ProvidencialData19

Datos por provincia proporcionados por ProvidencialData19

A partir de una [iniciativa lanzada por Pablo Rey](#) se reúnen un grupo de voluntarios que se hacen llamar [esCOVID19Data](#) (actualmente ProvidencialData19). Este grupo auto-



S

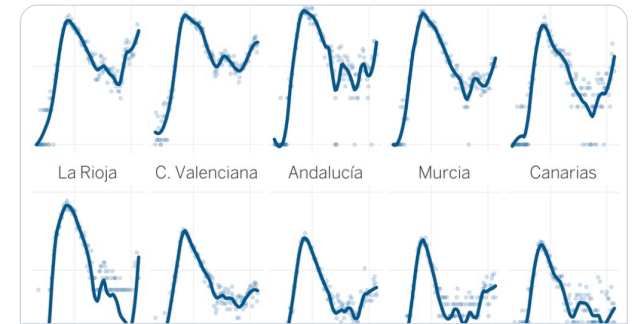


Instituto de Salud Carlos III

## Kiko Llaneras



El virus esta repuntando. Y lo peor es que parece generalizado: el aumento de contagios y hospitalizaciones se observa ya en diez comunidades. Repaso los datos oficiales con @borjandrinot y @danielegrasso:



El virus repunta en España: los datos de 10 comunidades muestran m...  
Análisis de las cifras oficiales en cada comunidad. No son comparables con marzo, porque la detección es mejor, pero el aumento de casos e



DATADISTA

Datadista



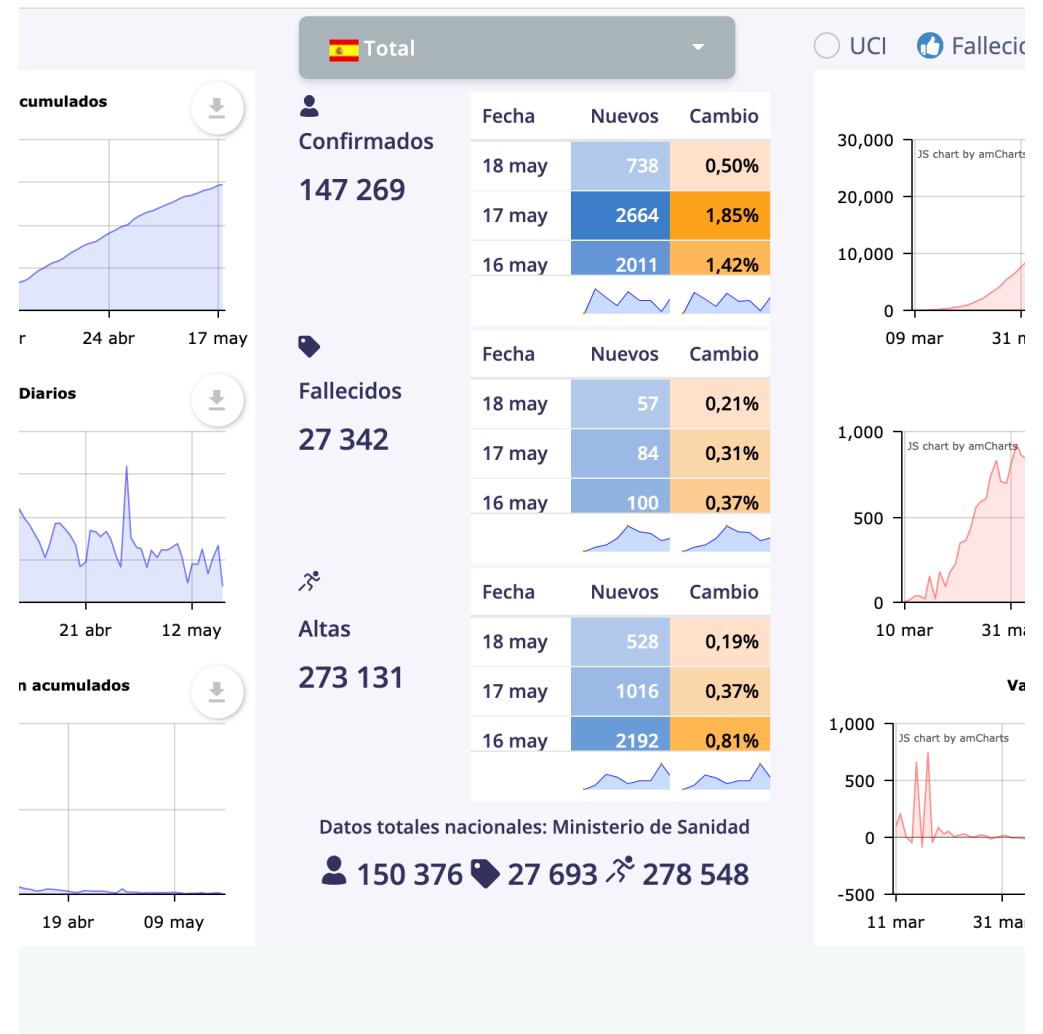
# CovidData-19

- **Nuestro portal ha permitido visualizar un problema matemático:**

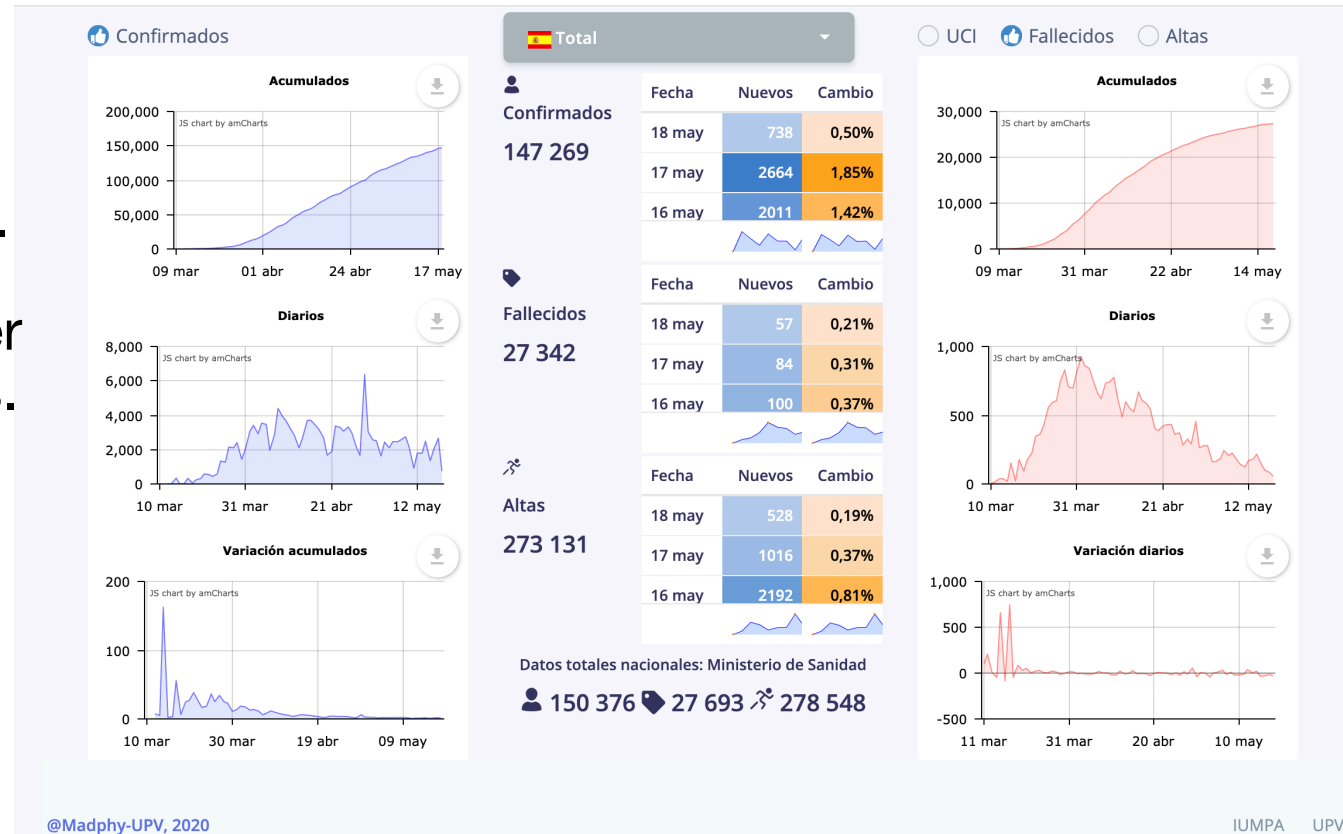
***la curva de supervivencia del virus***

- El portal ha sido creado por miembros del grupo de investigación
- Los datos han sido y son fundamentales para la gestión de la pandemia.
- Uno de los problemas proviene de que las comunidades tengan transferidas las competencias y al intentar aglutinar los datos es difícil sobre todo que se encuentren todos de manera sencilla y accesible.

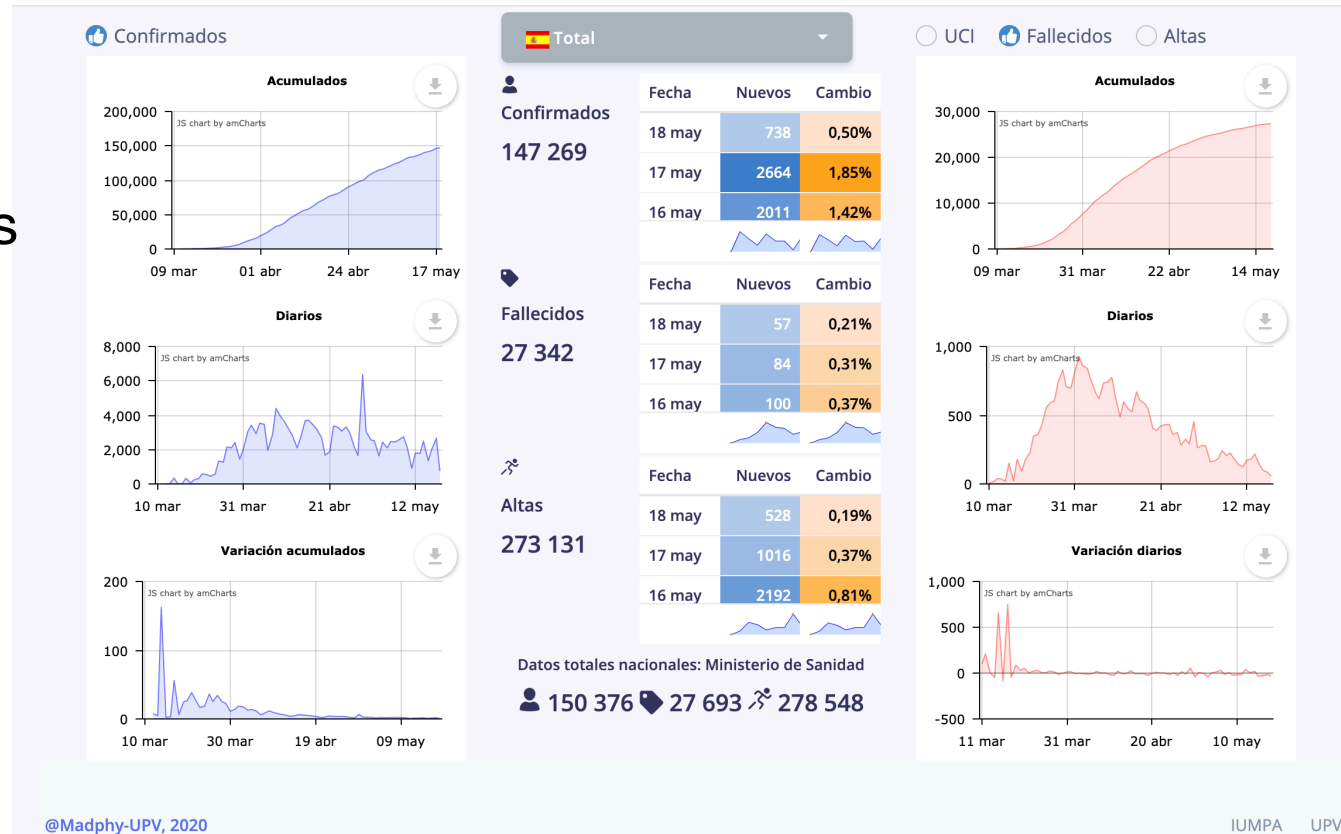
- La fuente de datos utilizada es la de Datadista.
- El portal tiene datos a nivel nacional.
- La primera página es un panel de control



- Así se encuentran los acumulados, la variación diaria.
- En el panel central se puede ver la evolución de los últimos días.
- Todo lo del portal se puede descargar y reutilizar.



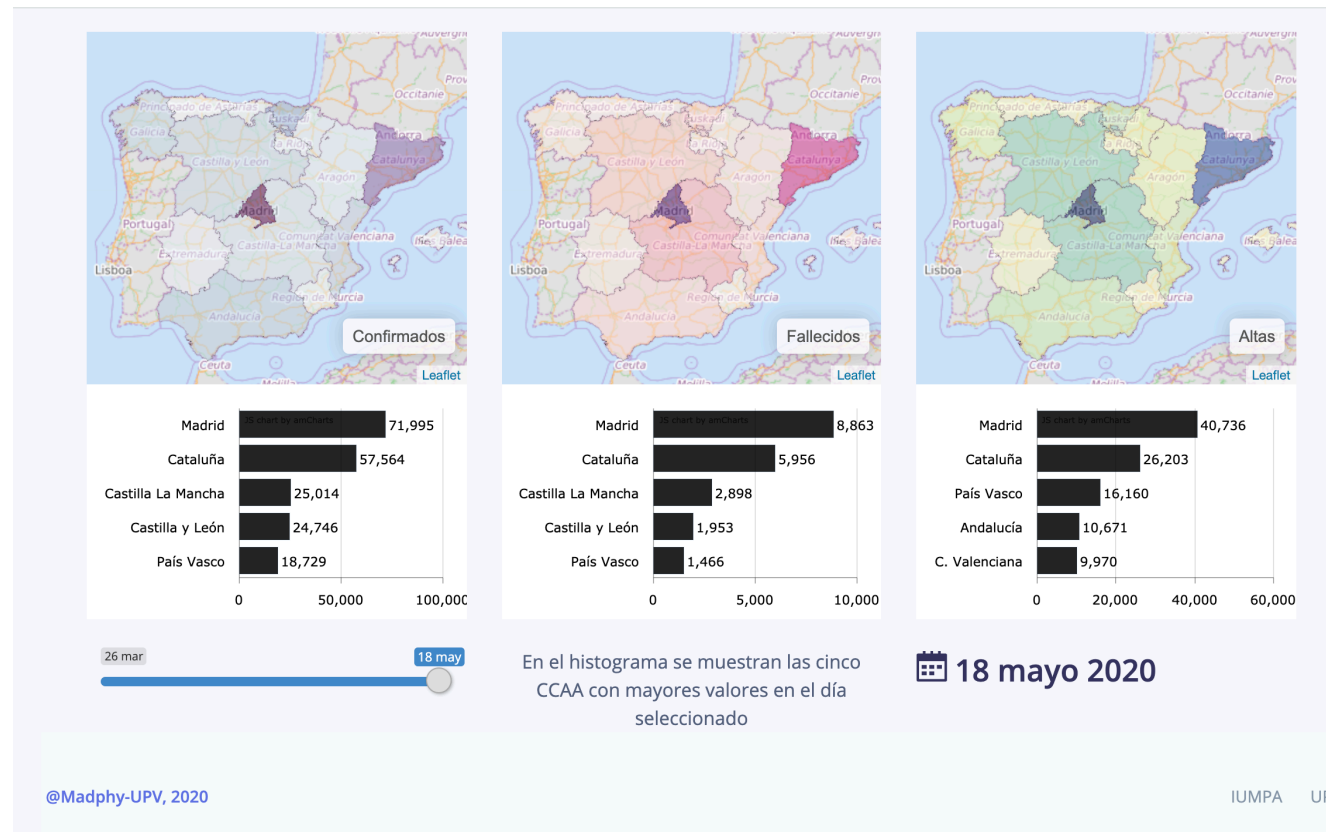
- En el periodo crítico, si se sumaban los desgloses de los datos de cada comunidad no sale el total nacional (uno de las mayores críticas hechas a Simón)
- Lo datos que nosotros publicamos: la suma de las tablas de datos que dan las comunidades al ministerio.
- El Ministerio da otra tabla distinta que son sus totales nacionales, ahí hay un desfase que se ha intentando corregir por parte del Ministerio.



Primer problema: No coincidencia de los datos nacionales con la suma de lo aportado por las CCAA

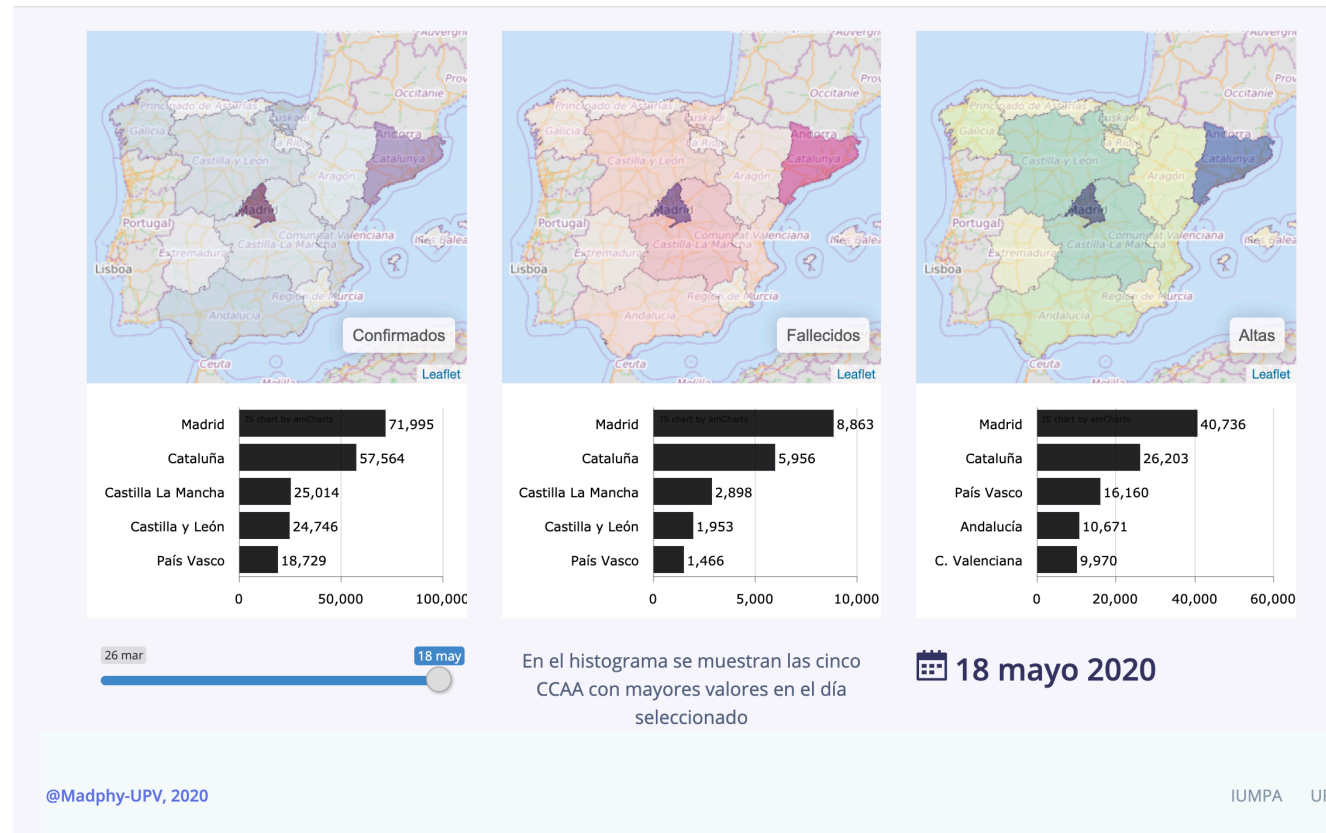


- La segunda pantalla aparece la representación de los datos en cada comunidad.
- Aparecen casos, variación diaria, y el porcentaje que representa del total.
- Se muestra una especie de ranking por comunidades.
- Otro de los problemas en el tratamiento de los datos han sido las variaciones a la hora de ofrecerlos. Así por ejemplo, al eliminar una columna de datos confirmados paso Ceuta a ser la mayor del ranking.

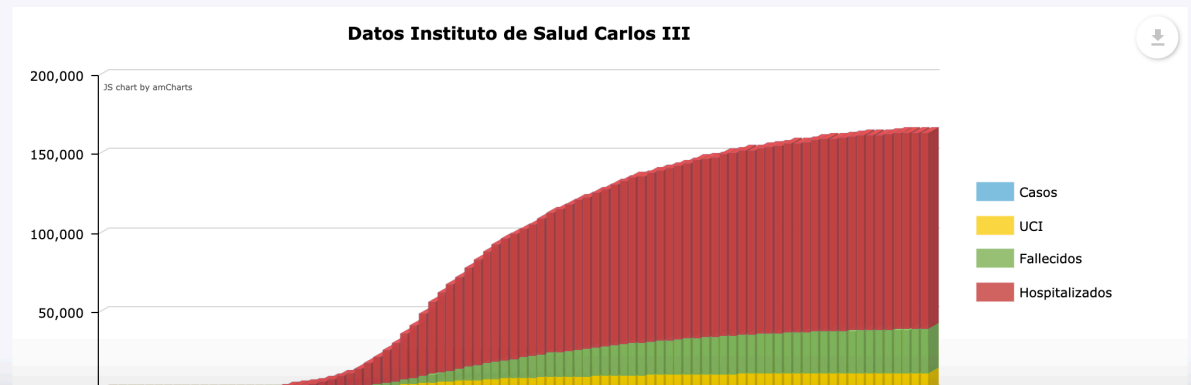
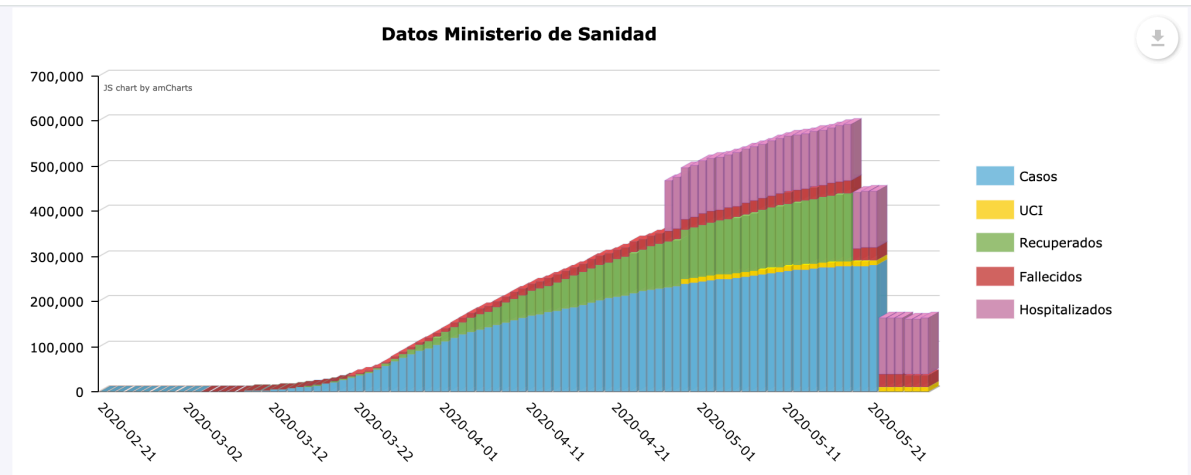


*Hay que estar siempre pendiente de que se ofrezcan siempre de la misma manera, cualquier cambio afecta, sobre todo si se esta automatizado.*

- El Ministerio de Sanidad no daba datos descargables en la crisis, se dan a través del Instituto de Salud Carlos III.
- Fue datadista la encargada de recoger los datos y son los que han utilizado la mayoría de medios de comunicación.
- Es un gran problema que no haya un organismo central encargado de la recogida de datos, lo que ha provocado que cada 4 o 5 días hubiera problemas a la hora de mostrar los datos.



- Mostramos los datos que da el Ministerio y los que da el Instituto de Salud Carlos III.
- Se ha incrementado la información del Instituto de Salud Carlos III si son fallecidos con PCR



Datos nacionales

Edad y sexo

Ranking

Otros datos

Dato nacional



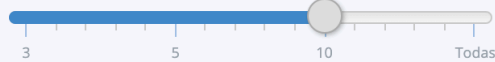
OFF

Número de comunidades

3

10

Todas



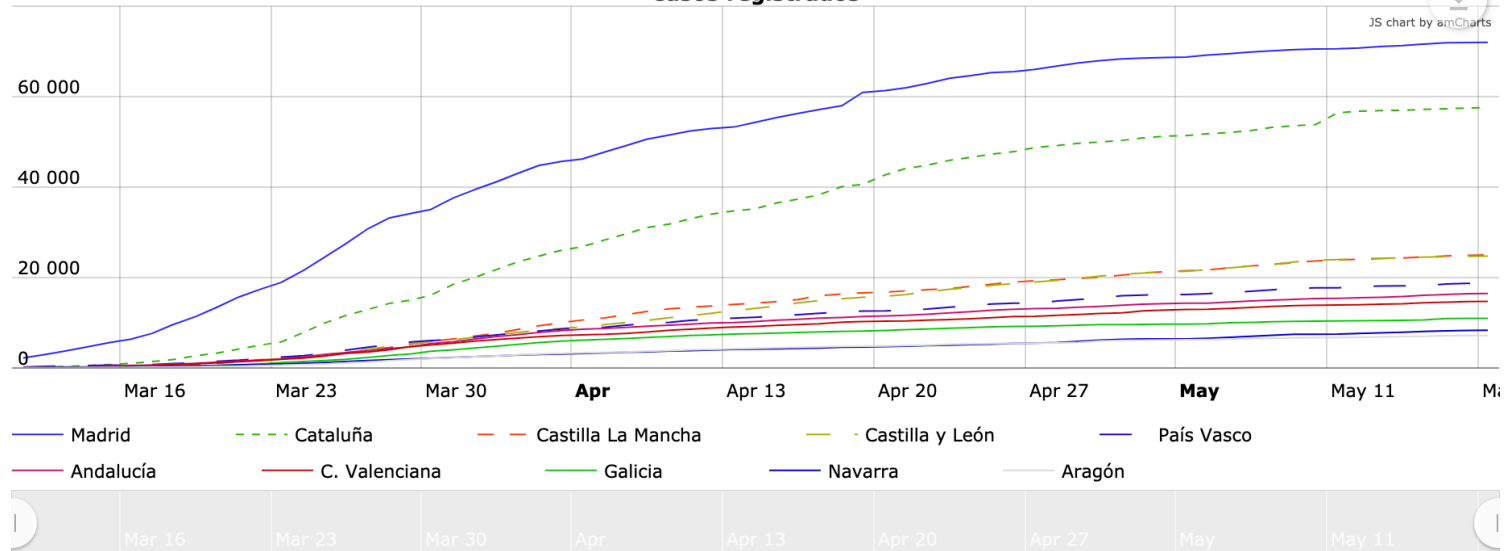
Confirmados

UCI

Altas

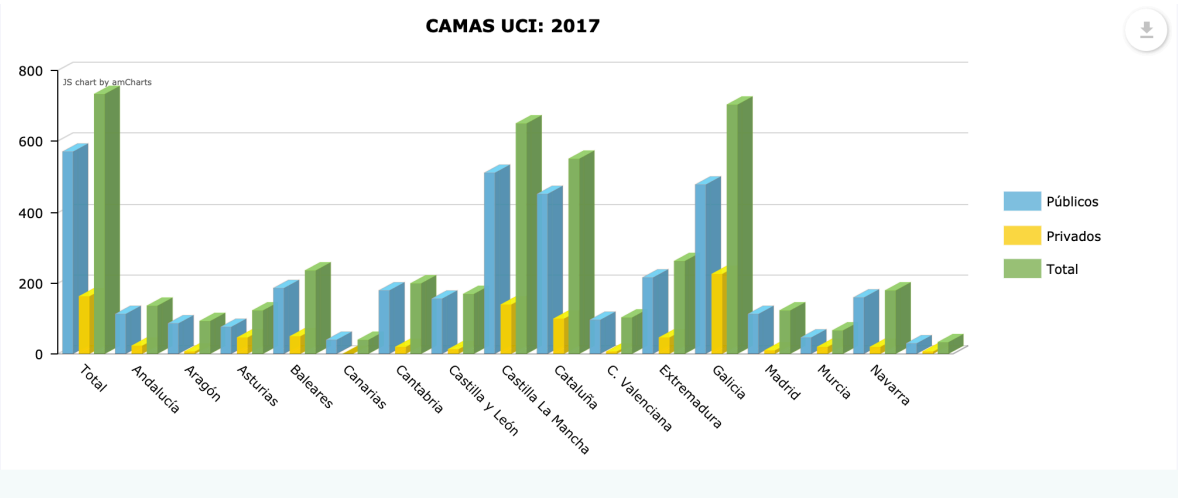
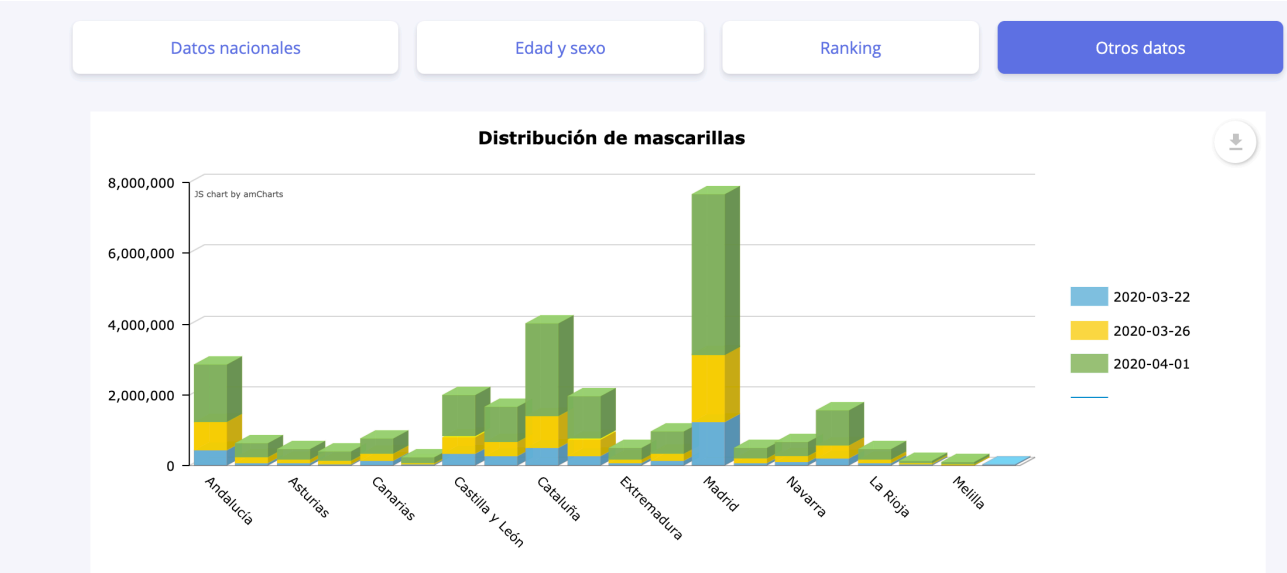
Fallecidos

Casos registrados



Zoom: All

A partir de los datos aportados por Datadista





# **Gestión para reducir la supervivencia de la Covid-19**

- Gestión de los datos: recuento, seguimiento... del número de incidencias
- Gestión clínica de la enfermedad
- Gestión política
- Datos demográficos, sociales...

**Todo ello como se ha realizado a nivel global, no nacional**

# Gestión para reducir la supervivencia de la Covid-19

- ¿Qué hacer con los datos?
  - En un primer momento el análisis de la famosa curva. CURVA DE CONTAGIOS.
    - Refleja cómo se expande la pandemia. Si crece mucho nos asustamos porque esta descontrolada la pandemia.
  - Otros aspectos a analizar desde el punto de vista estratégico: **Modelación de la curva de supervivencia.** *Esta curva analiza la gestión de las entradas y salidas en los centros sanitarios.*

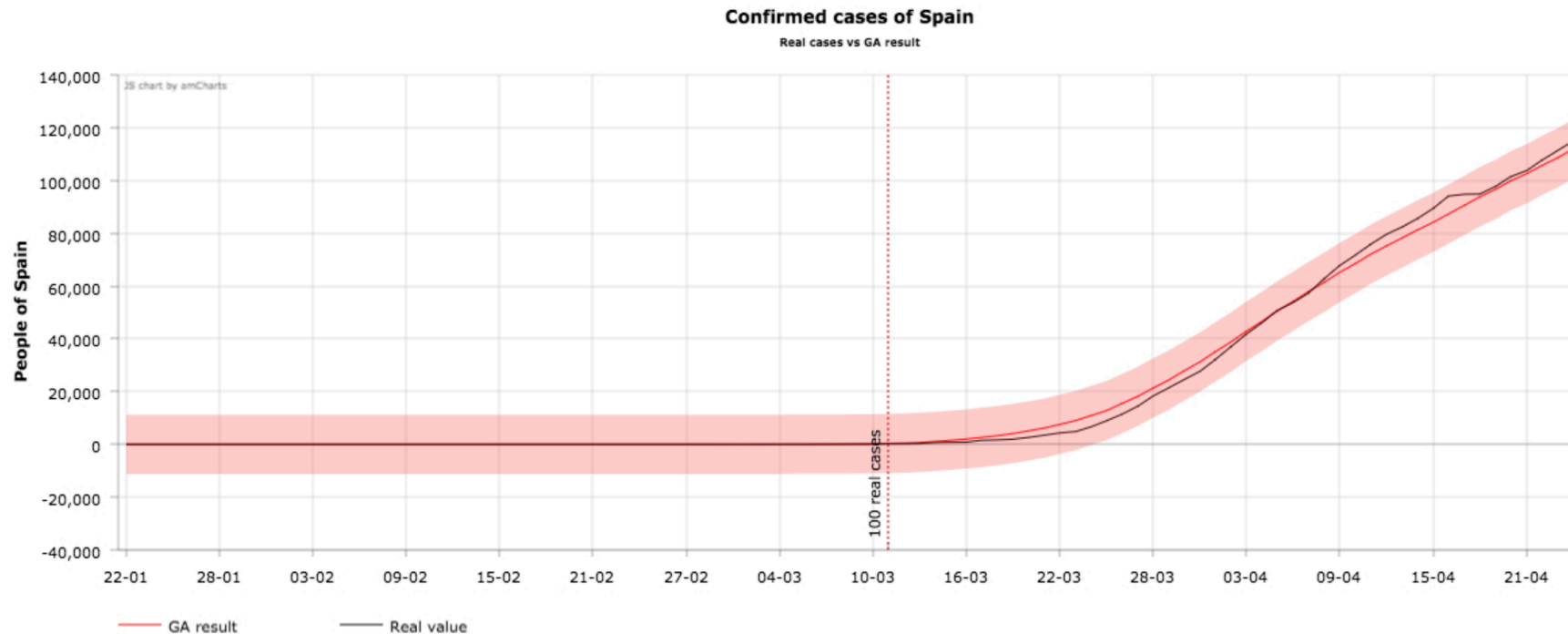
# Gestión para reducir la supervivencia de la Covid-19

- Las curvas muestran el número de pacientes infectados frente a las salidas del sistema: fallecidos y altas. Esto da unas curvas que muestran el tiempo medio de un individuo tutelado por el sistema sanitario.
- La curva de supervivencia en España. Indica la probabilidad de un individuo cuando entra en el sistema y cuanto tiempo puede tardar en salir.
- El problema para esta curva es no conocer exactamente el número de infectados. Esta curva muestra la capacidad del estado de detectar el número de infectados y sobre estos se cuantifica el número de fallecidos y altas.
- En Alemania a los diez días de entrada en el sistema cae en picado, los casos son más leves. Puede deberse al control de la población infectada.

# Gestión para reducir la supervivencia de la Covid-19

- La curva de supervivencia no tiene que ver con la de contagios.
- /Users/antoniaferrer/Google Drive/Politecnico/Articulos\_Congresos/2020/Mesa\_redonda\_ayuntamiento/Ficheros\_html/ComunitatValenciana\_Evolucio.html
- /Users/antoniaferrer/Google Drive/Politecnico/Articulos\_Congresos/2020/Mesa\_redonda\_ayuntamiento/Ficheros\_html/ComunitatValenciana.html

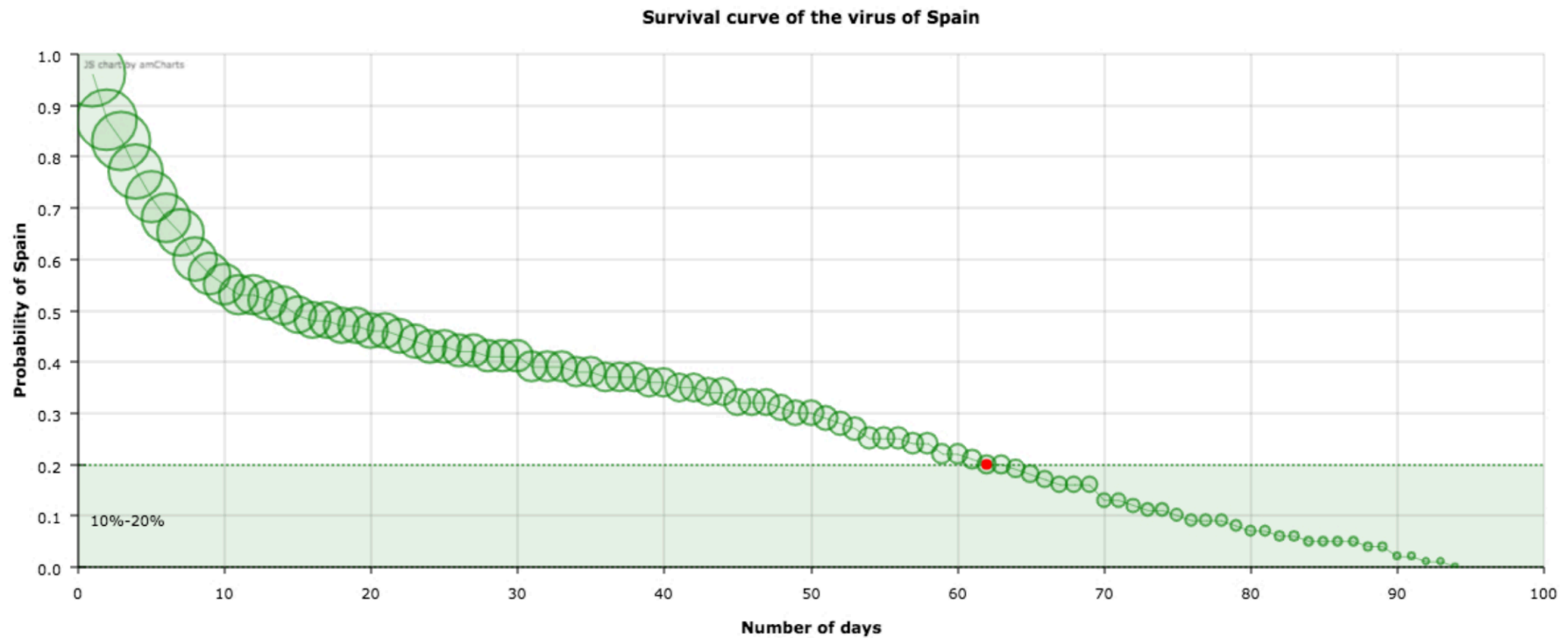
# Curvas de diseminación Kaplan-Meier



Curva de evolución de los casos confirmados (acumulados) para España: la línea negra corresponde al caso real, E, obtenido a partir de los datos y la línea roja es el resultado dado por el GA

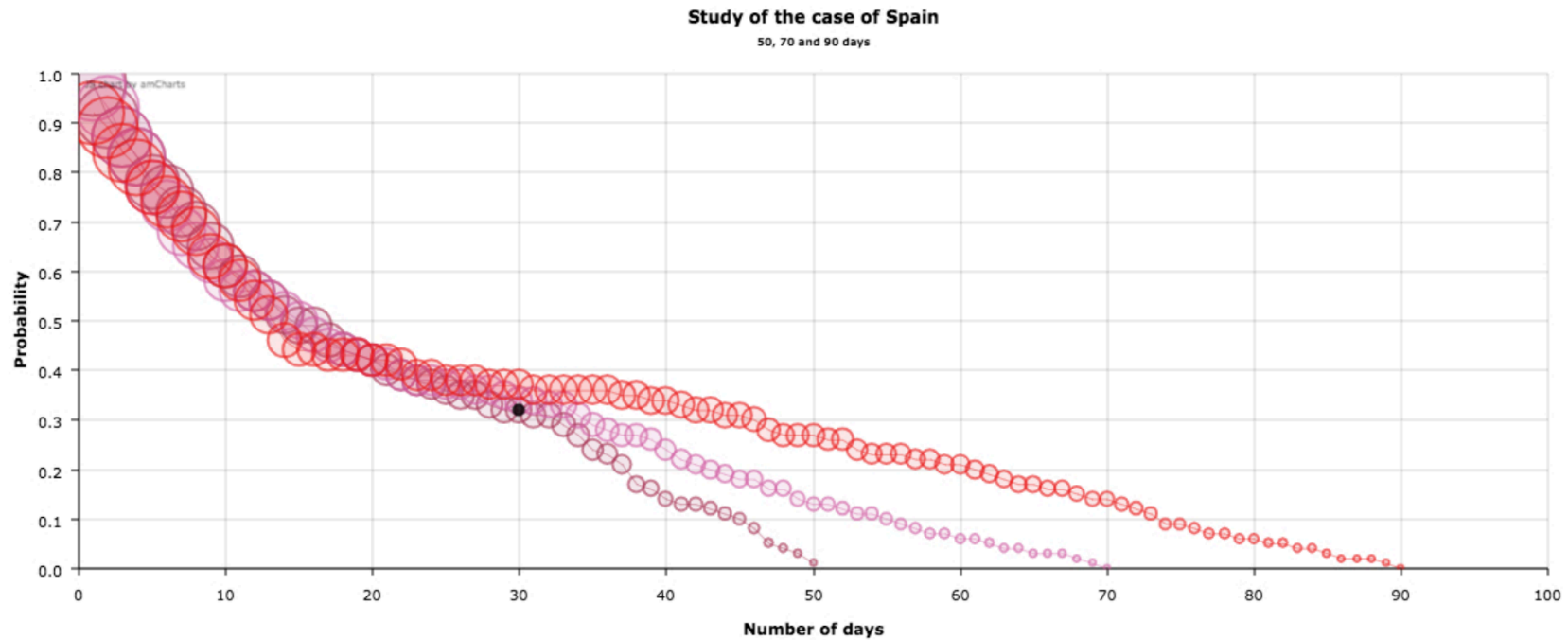


# Curvas de diseminación Kaplan-Meier



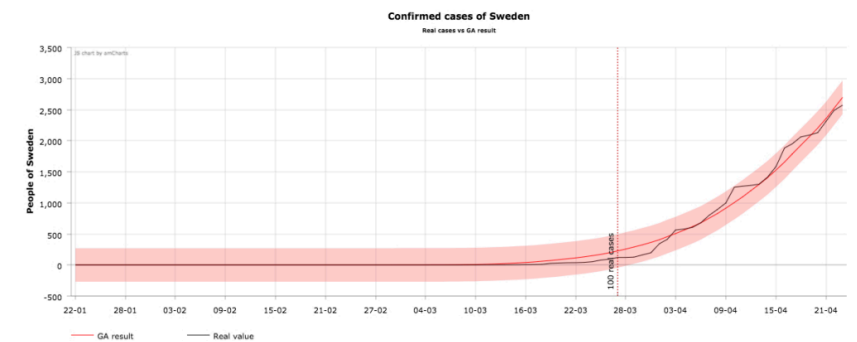
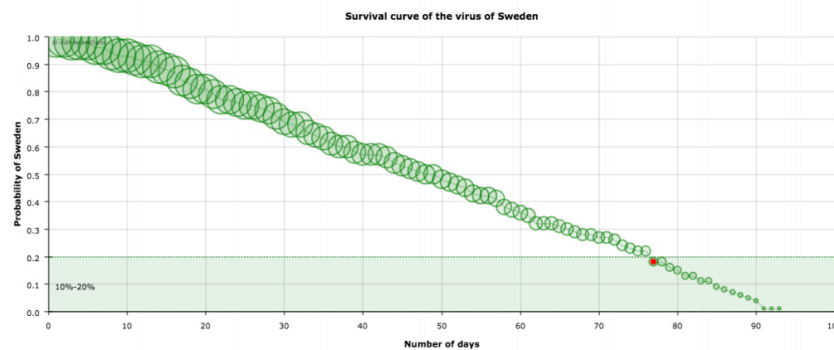
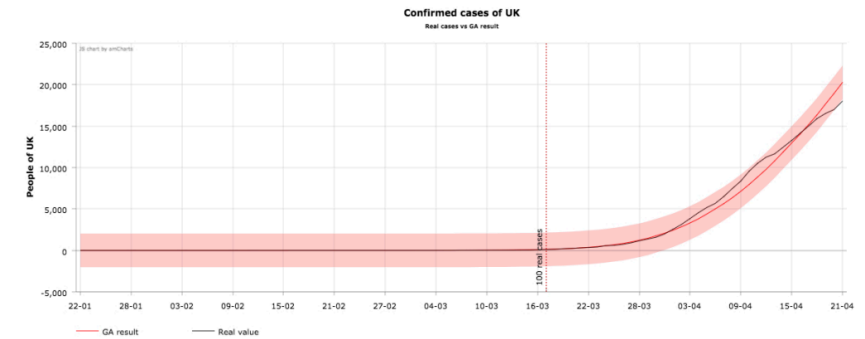
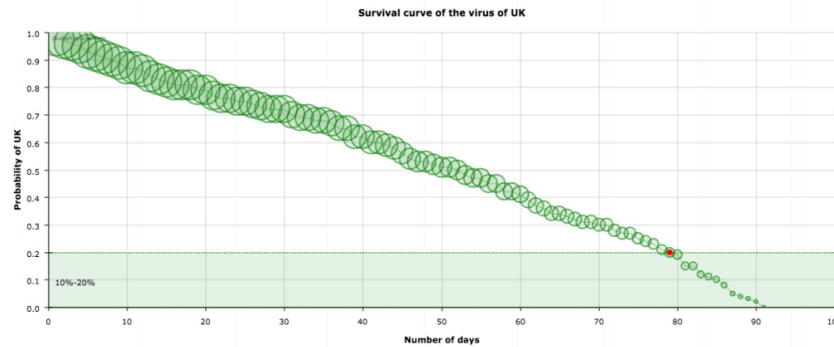
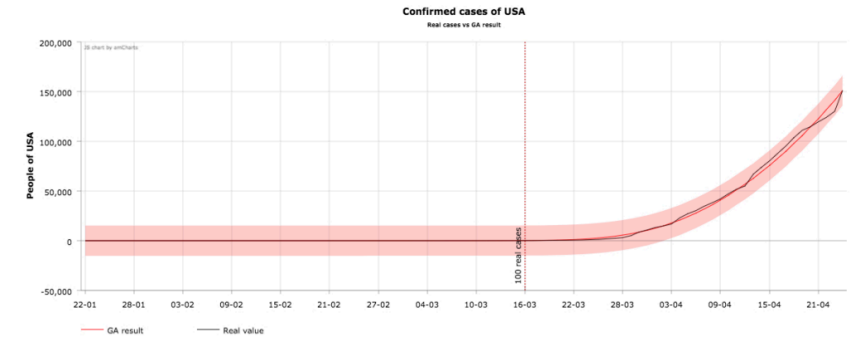
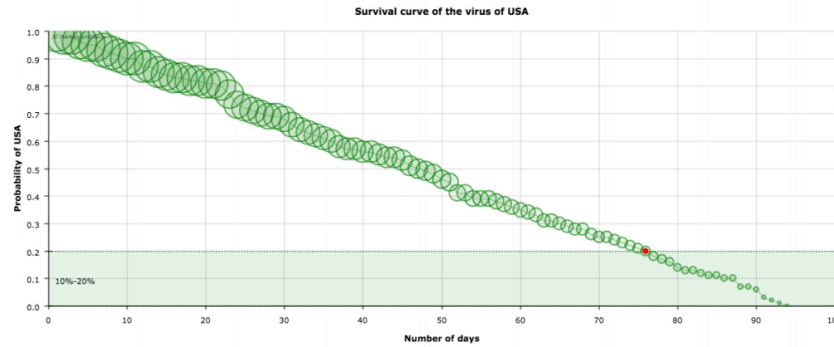
Curva de supervivencia, S, del virus correspondiente a España. El punto rojo es el día en que un individuo estándar tiene una probabilidad de salir del grupo de infectados inferior a 0,2

# Curvas de diseminación Kaplan-Meier



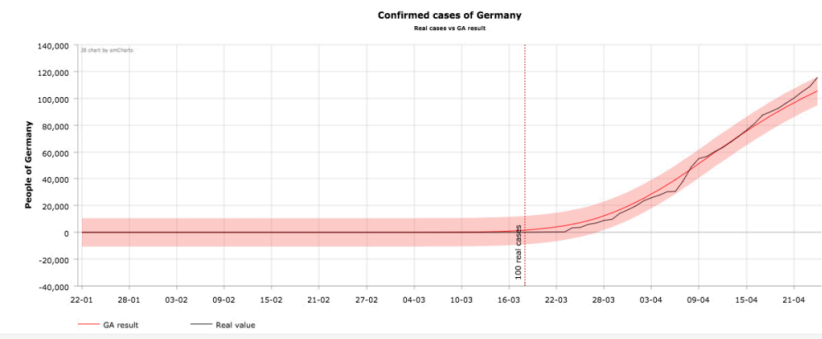
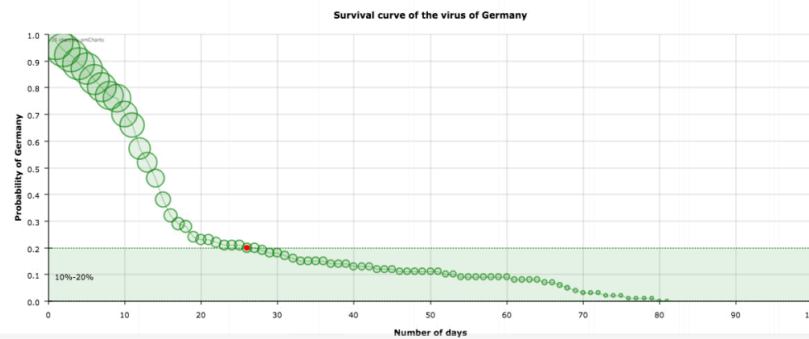
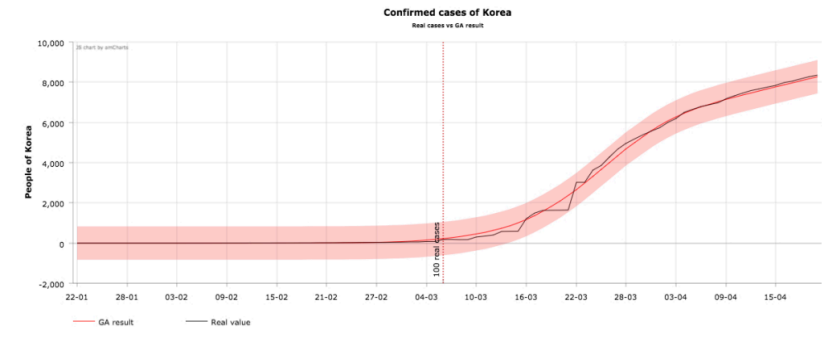
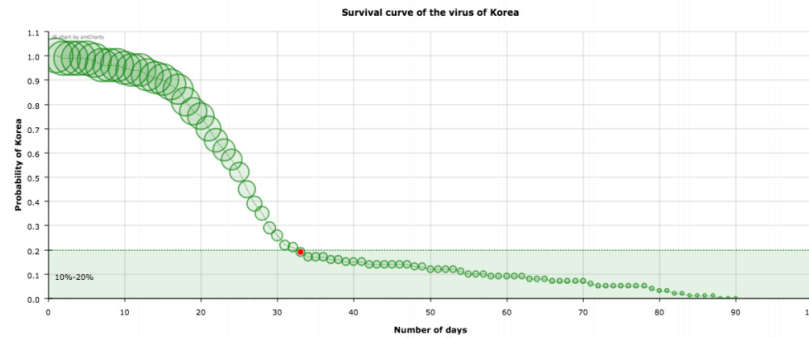
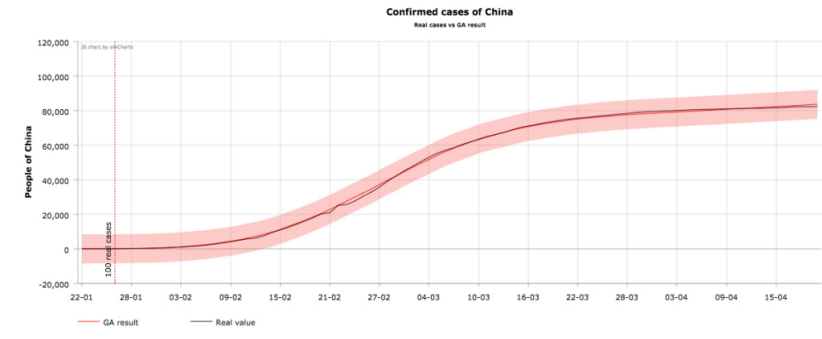
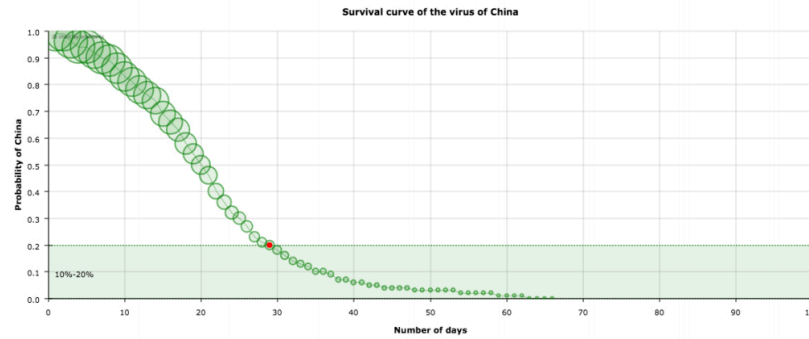
Curva de supervivencia del virus correspondiente a España para un número diferente de días. Los círculos rojos corresponden a los datos de nuestra serie temporal completa de días (90 días), los círculos rosados a 70 días y los círculos granate a 50 días.

# Curvas de diseminación Kaplan-Meier



Curvas de supervivencia del virus y curvas de casos confirmados (acumulados) correspondientes a EE.UU. (arriba), Reino Unido (centro) y Suecia (abajo)

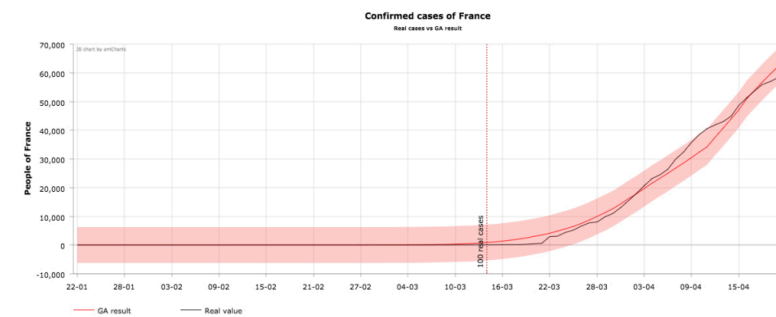
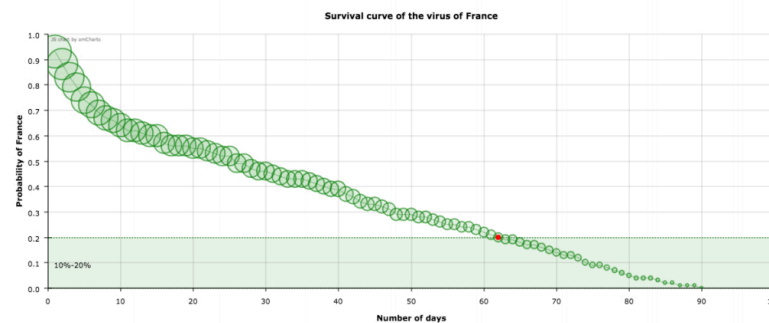
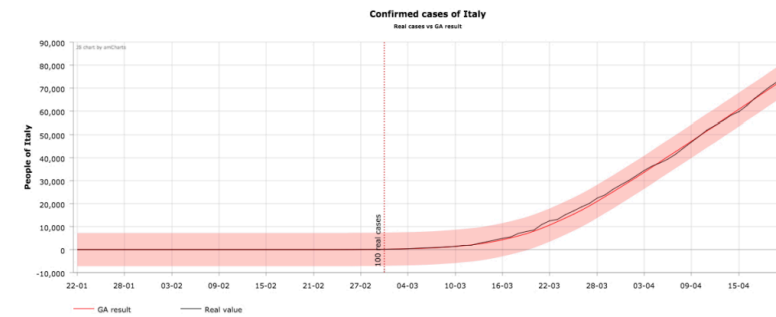
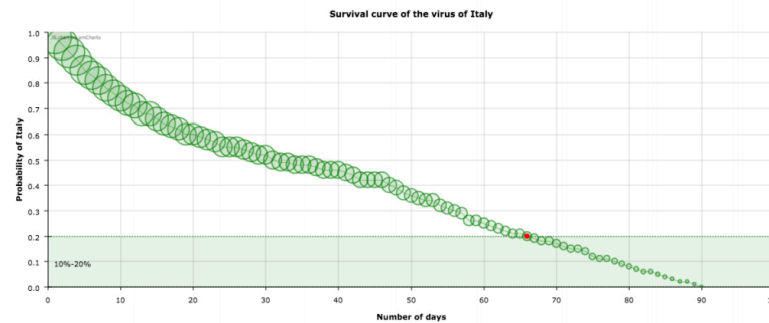
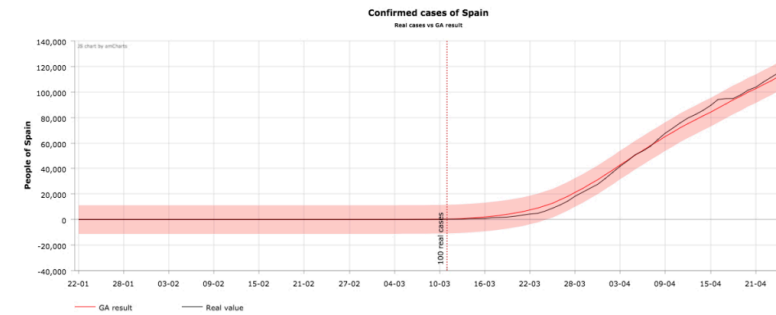
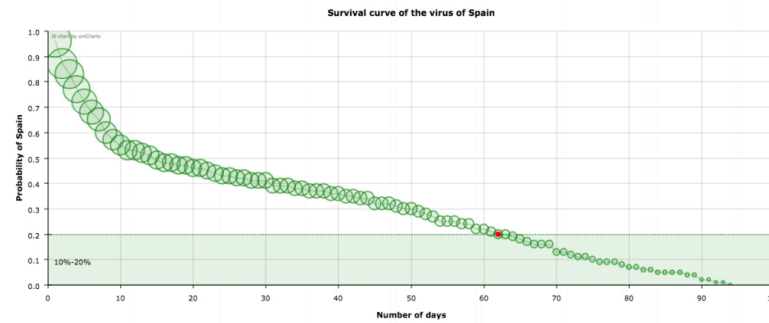
# Curvas de diseminación Kaplan-Meier



Curvas de supervivencia del virus y curvas de casos confirmados (acumulados) correspondientes a China (arriba), Corea del Sur (centro) y Alemania (abajo).

# Curvas de diseminación Kaplan-Meier

Survival curves of the virus and curves of (accumulated) confirmed cases corresponding to Spain (top), Italy (center) and France (bottom).





# Referencias

- Calabuig, J. M., García-Raffi, L. M., García-Valiente, A., & Sánchez-Pérez, E. A. (2020). Kaplan-Meier type survival curves for Covid-19: a health data based decision-making tool. arXiv preprint arXiv:2005.06032. <https://arxiv.org/pdf/2005.06032.pdf>
- Calabuig, J. M., García-Raffi, L. M., García-Valiente, A., & Sánchez-Pérez, E. A. (2020). Evolution model for epidemic diseases based on the Kaplan-Meier curve determination. Mathematics, 8 (8), 1260.



Imagen de misterfarmer en Pixabay



# Referencias

- Cobarsí-Morales, Josep (2020). “Covid-19: fuentes de información cuantitativa”. *Anuario ThinkEPI*, v. 14, e14d02.
- Ferrer-Sapena, Antonia; Calabuig, José-Manuel; Peset, Fernanda; Sánchez-del-Toro, Isabel (2020). “Trabajar con datos abiertos en tiempos de pandemia: uso de covidDATA-19”. *Profesional de la información*, v. 29, n. 4, e290421. <https://doi.org/10.3145/epi.2020.jul.21>



Imagen de misterfarmer en Pixabay

# ¡GRACIAS!